طراحی یک سیستم کنترل کیفیت چندمتغیره برای فرایندهای تولید چندمرحله‌یی

سیدتقی اخوان‌نیاکی (دانشیار)
دانشکده‌ی مهندسی صنایع، دانشگاه صنعتی شریف

در بسیاری از واحدهای تولیدی، که مراحل تولید محصول در آنها متوالی است، شاخص‌های کیفی محصول مستقل از هم نیست. به علاوه، در هر یک از این مراحل شاخص‌های کیفی را نیز نمی‌توان مستقل از هم فرض کرد. این تحقیق به شناسایی مرحله و عامل مؤثر در خارج از کنترل بودن یک فرایند تولیدی که در چند مرحله‌ی متوالی انجام می‌شود و در هر مرحله شاخص‌های کیفی وابسته وجود دارد، اختصاص دارد. بر این اساس، از روشی به‌نام «بخش‌بندی فضای پیشامدها» برای مشخص کردن مراحل مظنون استفاده شده است که در آن محدودسازی فضای جواب و حذف شاخص‌های با احتمال انحراف کمتر مطرح است. در تعیین متغیر یا متغیرهایی که در مرحله‌ی خارج از کنترل سبب انحراف شده‌اند، ترکیبی از روش‌های «تحلیل مؤلفه‌های اصلی» و «تحلیل تبعیض گذاری» به‌کار گرفته شده است.

مقدمه

انسان از دیرباز، از آن هنگام که به ساخت اولین محصول پرداخت، همواره برای ارتقاء کیفیت محصولاتش تلاش کرده است. ظرافت، دقت و هنرمندی خاصی که در آثار باستانی و محصولات مربوط به زمان‌های نه‌چندان دور مشاهده می‌شود حکایت از این تلاش دارد. با شروع انقلاب صنعتی ابزار و مهارت فردی هنرمندان و صنعتگران به‌تدریج جای خود را به ماشین‌های تولیدی سپرد و با پیدایش روش‌های جدید و پیچیده‌ی تولید، نیاز به کنترل کیفیت محصول نهایی ابعاد تازه‌یی یافت.

با توجه به این‌که تغییرپذیری پدیده‌یی دائمی، و جزء لاینفک همه‌ی محصولات تولیدی است، روش‌های آماری می‌توانند مؤثرترین وسیله برای بررسی و کنترل این تغییرات باشند. به طور کلی، روش‌های آماری بیشتر در کنترل کیفیت محصولاتی مؤثرند که به‌صورت انبوه تولید می‌شوند. نوع خاصی از مسائل کنترل کیفیت در شرایطی مطرح است که در آن چند شاخص کیفی به طور همزمان مورد توجه باشد. در این حالت، اگر شاخص‌ها متغیرهای تصادفی مستقل از هم نباشند، مسائل از پیچیدگی‌های بیشتری نسبت به مسائل کنترل کیفیت یک متغیره برخوردارند.

متأسفانه استفاده از چارت‌های کنترلی یک‌متغیره برای کنترل جداگانه‌ی شاخص‌های وابسته و مؤثر در کیفیت، یکی از روش‌های مرسوم در کنترل کیفیت به شمار می‌آید. در این حالت، عدم توجه به همبستگی بین شاخص‌های کنترلی و اعمال کنترل‌های جداگانه برای هر شاخص کیفی، ممکن است از نظر کیفیت محصول نهایی عواقب ناخوشایندی در بر داشته باشد. برای دوری از این کاستی، امروزه روش‌های آماری متعددی که در آنها همبستگی بین شاخص‌های کنترلی ملحوظ است در اختیارند.

در کنترل کیفیت چندمتغیره، برای تعیین حالت خروج از کنترل روش‌های گوناگونی وجود دارد. تقریباً در همه‌ی این روش‌ها فرض نرمال چندمتغیره حاکم است، در حالی‌که در برخی از آنها دسترسی به پارامترهای جامعه‌ی آماری، یعنی بردار میانگین و ماتریس واریانس ـ کوواریانس، پیش‌نیاز است و در برخی دیگر این پارامترها براساس نمونه‌های تصادفی مناسب از فرایندی که تحت کنترل است، برآورد می‌شود.

برای تعیین وضعیت کلی فرایند در هر زمان، آزمون فرضی مبنی بر برابری بردار میانگین فرایند در آن زمان با بردار میانگین در دست یا برآورد شده، انجام می‌شود که در آن فرض می‌شود ماتریس‌های واریانس ـ کوواریانس جمعیت مرجع و جدید برابرند. تاکنون آزمون‌های مختلفی برای برابری ماتریس‌ها ارائه شده است.[۱و۲] اگر جامعه‌ی مورد نظر از توزیع نرمال چندمتغیره برخوردار باشد و ماتریس‌های واریانس ـ کوواریانس جمعیت‌های مرجع و جدید آن برابر باشند می‌توان از آماره‌ی هاتلینگ[۳] برای خارج از کنترل بودن کلی فرایند استفاده کرد. ولی این روش راهی برای تعیین متغیر یا متغیرهایی که خارج از کنترل بودن فرایند را سبب شده‌اند، فراهم نمی‌کند.

برای کنترل جداگانه‌ی متغیرهای موجود در فرایند آلت روشی تقریبی را براساس مفهوم نامساوی بونفرونی[۱] ارائه داده است[۳] و

دیگران[۲] نیز کارایی این روش را مورد بررسی قرار داده و به این نتیجه رسیده‌اند که علی‌رغم مؤثر بودن این روش، در برخی موارد تنگ‌تر شدن فواصل اطمینان منجر به ظهور علائم خروج از کنترل فرایند، در حین تحت کنترل بودن فرایند می‌شود.

محققان با انتخاب کلیه‌ی زیرمجموعه‌های متشکل از شاخص‌های کیفی موجود در فرایند و محاسبه‌ی آماره‌ی هاتلینگ برای هریک از آنها، روش‌هایی براساس بخش‌بندی فضای پیشامدها ارائه داده‌اند.[۴و۵] کارایی این روش‌ها با افزایش تعداد شاخص‌های مورد بررسی، به‌سبب امکان وجود حالت‌های فراوان برای زیرمجموعه‌ها، کاهش چشمگیری می‌یابد.

در تعیین علل خارج از کنترل بودن فرایند، محققان مفهوم تجزیه و تحلیل مؤلفه‌های اصلی را به‌کار گرفتند.[۳و۶] در این روش آماره‌ی هاتلینگ به مجموع مربعات مؤلفه‌های اصلی تجزیه می‌شود که هر یک از ویژگی استقلال برخوردارند و هریک بیان کننده‌ی یک بخش یا تمامی تغییرات ناشی از متغیرهای اولیه‌اند. به همین دلیل نمی‌توان به‌سادگی تغییر ناشی از تبدیلی راکه در مفهوم تحلیل مؤلفه‌های اصلی مطرح است به متغیر عامل انحراف اولیه نسبت داد.

دوگاناکسوی و همکاران[۳] یک روش رتبه‌بندی را بر اساس آماره‌ی یک متغیره‌ی t استودنت ارائه داده‌اند. بدین ترتیب که اگر تغییر قابل ملاحظه‌یی در مقدار آماره‌ی هاتلینگ مربوط به مجموعه‌یی از شاخص‌های کیفی مشاهده شد، آنگاه بیشترین احتمال مربوط به شاخصی است که دارای بزرگ‌ترین مقدار آماره‌ی t است. آنها به‌منظور دریافت علامت خروج از کنترل برای یکایک شاخص‌ها، از فواصل اطمینان نوع بونفرونی[۲] استفاده کردند. یکی از مشکلات این روش تعیین کوچک‌ترین سطح اطمینان برای هر شاخص است.

مندال در سال ۱۹۶۹ روش رگرسیون را برای اولین بار در نمودارهای کنترل کیفیت به‌کار گرفت.[۷] این ایده بعدها در تهیه‌ی چارت‌های کنترلی انتخاب عامل خارج از کنترل مورد استفاده قرار گرفت.[۸] هاوکینز ایده‌ی مندال و نتایج کار هیلی را با یکدیگر تلفیق کرد.[۹-۱۱] بدین ترتیب که برای هر متغیر یک معادله‌ی رگرسیون متشکل از سایر متغیرها تعیین کرد و از طریق شناسایی متغیری که بیشترین اثر را در معادله‌ی رگرسیون منعکس می‌کند، متغیری راکه عامل انحراف است شناسایی کرد. یکی از مشکلات این روش، عدم دسترسی به یک رابطه‌ی خطیِ گویا برای بیان تمامی تغییرات ناشی از انحراف هر متغیر در متغیر وابسته‌ی موجود در معادله‌ی رگرسیون است.

در این تحقیق با بهره گرفتن از ایده‌هایی که در روش‌های بخش‌بندی پیشامدها، مؤلفه‌های اصلی و رتبه‌بندی براساس آماره‌ی t استودنت وجود دارد، یک روش ترکیبی مورد بررسی قرار می‌گیرد.

دیدگاه کلی مراحل تحقیق

فرض بر این است که در هر یک از مراحل، فرایند تولید از یک توزیع نرمال چندمتغیره برخوردار است. این توزیع از طریق یک بردار میانگین و یک ماتریس واریانس ـ کوواریانس کاملاً شناخته می‌شود. ولی در عمل معمولاً بردار میانگین و نیز ماتریس واریانس ـ کوواریانس این بردار تصادفی از پیش مشخص نیست و باید آنها را از طریق یک فرایند تحت کنترل تخمین زد. برای این‌کار، در نخستین مرحله که مرحله‌ی راه‌اندازی نام دارد، باید با استفاده از روش آلت[۱۲] و از طریق بردار میانگین نمونه و ماتریس واریانس ـ کوواریانس نمونه، بردار میانگین و ماتریس واریانس ـ کوواریانس را تخمین زد. پس از تخمین و محاسبه‌ی مقادیر مورد نظر در مرحله‌ی بعد که مرحله‌ی بررسی کل فرایند از جهت انطباق با شرایط تحت کنترل است با استفاده از آماره‌ی هاتلینگ، اگر فرایند خارج از کنترل باشد، مراحل پدید آورنده‌ی کل فرایند مورد بررسی قرار می‌گیرند. این مرحله به‌عنوان یک مرحله‌ی میانبر برای شناسایی عوامل انحراف کل فرایند از شرایط تحت کنترل تلقی می‌شود (در این خصوص به تفصیل روش جدیدی ارائه می‌شود). با تعیین مرحله یا مراحل عامل انحراف، تعیین متغیر یا متغیرهای عامل انحراف که هدف نهایی است، در مرحله‌ی بعدی مورد بررسی قرار می‌گیرد. در این مرحله، با استفاده از روش مؤلفه‌های اصلی سعی بر آن است که هرچه بیشتر متغیر یا متغیرهای مظنون شناسایی شوند. برای هرچه بهتر نشان دادن کاربرد مؤلفه‌های اصلی در مبحث کنترل کیفیت چند متغیره، روشی مرکب از این شیوه و شیوه‌ی انتخاب زیرمجموعه‌یی از متغیرها ارائه خواهد شد. به‌طور خلاصه، ترتیب موارد مورد بحث در این پروژه چنین است:

۱. مرحله‌ی راه‌اندازی؛
۲. تعیین وضعیت کل فرایند از نظر انطباق با شرایط تحت کنترل؛
۳. تعیین مرحله یا مراحل مظنون به عامل انحراف فرایند از شرایط تحت کنترل؛
۴. تعیین متغیر یا متغیرهای مظنون به عامل انحراف کل فرایند از شرایط تحت کنترل؛

مرحله‌ی راه‌اندازی

فرایندی رادر نظر بگیرید که از k مرحله تشکیل شده است. فرض کنید که در هر مرحله‌ی تولیدی تعداد p شاخص کیفیتی وابسته وجود

دارد که از یک توزیع نرمال p متغیره با بردار میانگین μ و ماتریس واریانس ـ کوواریانس Σ برخوردارند (یعنی متغیرهای کیفی دارای توزیع $N_p(\mu,\Sigma)$ هستند). در مرحله‌ی راه‌اندازی m زیرگروه از داده‌ها موجود است و در هر زیرگروه n مشاهده برای p متغیر در نظر گرفته می‌شود. نمادهایی که در بخش‌های آتی به کار خواهند آمد عبارتند از:

مقدار مشاهده شده‌ی lام از متغیر iام در زیرگروه jام: x_{ijl}

میانگین نمونه‌ی متغیر iام در زیرگروه jام: $\bar{x}_{ij}$

$$\bar{x}_{ij}=\frac{\sum_{l=1}^{n} x_{ijl}}{n} \quad , \quad (i=1,2,...,p \;\; ; \;\; j=1,2,...,m) \qquad (1)$$

میانگین نمونه‌ی متغیر iام:

$$\bar{\bar{x}}_{ij}=\frac{\sum_{j=1}^{m} \bar{x}_{ij}}{m} \quad ; \quad (i=1,2,...,p) \qquad (2)$$

بردار میانگین نمونه:

$$\bar{\bar{X}}=[\bar{\bar{x}}_1,\bar{\bar{x}}_2,...,\bar{\bar{x}}_p]^T \qquad (3)$$

برآورد عنصر سطر iام و ستون jام ماتریس واریانس ـ کوواریانس نمونه‌ی زیرگروه jام: S_{ij}^{j}

$$S_{ij}^{j}=\frac{1}{m-1}\sum_{l=1}^{m}(\bar{x}_{il}-\bar{\bar{x}}_i)(\bar{x}_{jl}-\bar{\bar{x}}_j) \; ; \; i,j=1,2,...,p \qquad (4)$$

ماتریس واریانس ـ کوواریانس میانگین نمونه:

$$S_p=\frac{\sum_{j=1}^{m} S^j}{m} \qquad (5)$$

معکوس ماتریس واریانس ـ کوواریانس میانگین نمونه: S_p^{-1}

مشاهدات مورد استفاده در زیرگروه‌ها که در تخمین به کار رفته‌اند باید از یک فرایند تحت کنترل آمده باشند. برای بررسی این مورد، در ادامه‌ی مرحله‌ی راه‌اندازی از آماره‌ی هاتلینگ به‌صورت:[۱۲]

$$T_j^2=n(\bar{\bar{X}}^{(j)}-\bar{\bar{x}})S_p^{-1}(\bar{\bar{X}}^{(j)}-\bar{\bar{x}}) \qquad (6)$$

استفاده می‌شود که در آن $\bar{\bar{X}}^{(j)}$ برداری p بعدی و شامل میانگین نمونه‌ی زیرگروه‌های مربوط به هر یک از متغیرهاست. مقدار این آماره با حد بالای کنترل که از رابطه‌ی زیر به‌دست می‌آید مقایسه می‌شود. در صورتی که مقدار T_j^2 مربوط به هر زیرگروه از حد بالای

$$UCL=\left(\frac{mnp-mp-np-p}{mn-m-p+1}\right)F_{\alpha,(p,mn-m-p+1)} \qquad (7)$$

کنترل بیشتر شود، باید مشاهدات آن زیرگروه، برای تشخیص عوامل قابل تشخیص، مورد بررسی قرار گیرد. در صورت یافتن این گونه عوامل، مشاهده‌ی مربوط به آن در محاسبات بعدی کاربرد ندارد. در این حالت فرض کنید بردار میانگین نمونه‌ی جدید $\bar{x}^*$ و ماتریس واریانس ـ کوواریانس جدید S_p^* باشد. برای بررسی قابل قبول بودن زیرگروه‌های به کار رفته در محاسبات اخیر، آماره‌ی هاتلینگ به‌صورت:

$$T^2=n(\bar{x}^{(future)}-\bar{\bar{x}}^*)(S_p^*)^{-1}(\bar{x}^{(future)}-\bar{\bar{x}}) \qquad (8)$$

محاسبه و با حد بالای کنترل به صورت:

$$UCL=\left(\frac{p(m-a+1)(n-1)}{n(m-a)-m+a-p+1}\right)F_{\alpha,p,n(m-a)-m+a-p+1} \qquad (9)$$

مقایسه می‌شود. در رابطه‌ی ۹، a تعداد زیرگروه‌های حذف شده در مرحله‌ی قبلی کنترل زیرگروه‌ها را نشان می‌دهد.[۱۲]

تعیین وضعیت کل فرایند از نظر انطباق با شرایط تحت کنترل

وضعیت کل فرایند را می‌توان پس از تخمین مناسب جامعه‌ی آماری که در مرحله‌ی راه‌اندازی صورت گرفت، با استفاده از آماره‌ی هاتلینگ بررسی کرد. در اینجا نیز فرض می‌شود که مشاهدات مربوط به مشخصه‌های کیفی از توزیع نرمال چندمتغیره برخوردارند. ذکر این نکته ضروری است که قبل از انجام هرگونه پردازش بر روی داده‌های نمونه، چه در مرحله‌ی راه‌اندازی و چه در مرحله‌ی تعیین وضعیت کلی فرایند، باید از برخورداری این داده‌ها از توزیع نرمال چندمتغیره اطمینان نسبی کسب کرد.[۲]

تعیین مراحل مشکوک به عامل انحراف

پس از تعیین وضعیت کلی فرایند و دریافت علامت خارج از کنترل بودن فرایند ـ از طریق آماره‌ی هاتلینگ در مرحله‌ی قبلی ـ تعیین مرحله یا مراحلی از فرایند که متغیر یا متغیرهای آنها دلیل انحراف کل فرایند از تحت کنترل بودن هستند، مطرح می‌شود. در این مرحله از روشی جدید استفاده می‌شود.

فرض کنید که فرایند کلی شامل k مرحله است و تعداد r_i شاخص کیفیتی (متغیر) در مرحله‌ی iام وجود دارد به‌نحوی که:

$$\sum_{i=1}^{k} r_i=p \qquad (10)$$

با استفاده از ایده‌ی روش مورفی[۴] که برای تعیین متغیر عامل انحراف در یک مرحله ابداع شد، برای تعیین مرحله‌ی عامل انحراف، ابتدا فضای پیشامد K بعدی را به دو زیر فضای K_1 و K_2 افراز می‌کنیم به‌نحوی که $K_1+K_2=K$ در این‌صورت آماره‌ی مناسب برای بررسی معنی‌دار بودن زیرفضای K_1 عبارت است از

$$D=T_K^2-T_{K_1}^2 \qquad (11)$$

که تحت فرض صفر از یک توزیع کای دو[۳] با $K_۲=K-K_۱$ درجه‌ی آزادی برخوردار است.[۴] به عبارت دیگر در سطح اطمینان $(۱-\alpha)$، با افراز فضای K بعدی، محاسبه‌ی آماره‌ی $T^۲$ هاتلینگ برای این فضا و زیرفضای $K_۱$، و مقایسه‌ی کم‌ترین مقدار D، که از مقایسه‌ی مقادیر D مربوط به زیرفضاهای با درجه‌ی آزادی یکسان به دست می‌آید، با مقدار بحرانی $\chi^۲_{\alpha,K_۲}$ زیرفضای مظنون به عامل انحراف مشخص می‌شود. در روش مورفی، افراز فضای K بعدی با این بینش انجام شده است که زیرفضایی که با کم‌ترین درجه‌ی آزادی از بیشترین مقدار $T^۲_{K_۱}$ برخوردار باشد، بیشترین اثرگذاری را در $T^۲_K$ خواهد داشت. ولی از آنجاکه در این تحقیق مراحل نقش متغیرها را بازی می‌کنند، نمی‌توان از این ایده برای تعیین زیرمجموعه‌های عامل انحراف استفاده کرد. برای این کار ضروری است که ابتدا فهرستی از ترکیبات مختلف درجه‌ی آزادی را برای زیرفضای شامل $K_۱$ درجه‌ی آزادی به دست آورد و آنها را به‌ترتیب صعودی مرتب کرد. اگر رابطه‌ی

$$r_۱ \le r_۲ \le ... \le r_K \qquad (۱۲)$$

برقرار باشد، این ترتیب به‌صورت جدول ۱ خواهد بود.

بعد از به‌دست آوردن جدول ۱، زیرمجموعه‌هایی که از درجه‌ی آزادی یکسانی برخوردارند در یک گروه قرار می‌گیرند. در این حالت با فرض دستیابی به f گروه، الگوریتم این تحقیق برای تعیین مراحل مظنون به عامل انحراف حداکثر دارای f گام خواهد بود که در هر گام گروه g_i مورد بررسی قرار می‌گیرد. البته انجام بررسی در گام i ام بستگی به نتیجه‌ی بررسی در گام $i-۱$ دارد. بدین صورت که فقط هنگامی به گام i ام می‌رویم که نتیجه‌ی بررسی در قدم $i-۱$ در تعیین

جدول ۱. ترتیب صعودی درجه‌ی آزادی برای زیرفضای شامل $K_۱$ درجه‌ی آزادی.

تعداد درجات آزادی باقی‌مانده ($K_۲$)	زیرفضای موردبررسی ($K_۱$)	زیرمجموعهٔ انتخابی از مراحل
$(K-r_۱)=r_۲+...+r_K$	$r_۱$	{۱}
$(K-r_۲)=r_۳+...+r_K$	$r_۲$	{۲}
.	.	
.	.	
.	.	
$(K-r_K)=r_۱+...+r_{K-۱}$	r_K	{K}
$(K-r_۱-r_۲)=r_۳+...+r_K$	$r_۱+r_۲$	{۱,۲}
.	.	
.	.	
.	.	
$r_۱$	$(r_۲+r_۳+...+r_K)$	{۲,۳,...K}

زیرمجموعه‌ی مظنون به عامل انحراف با شکست مواجه شده باشد. در هر گام کم‌ترین مقدار D، که با نماد D_s نشان داده می‌شود، با مقدار بحرانی $\chi^۲_{\alpha,K_۲}$ مقایسه می‌شود و اگر شرط $D_s \le \chi^۲_{\alpha,K_۲}$ برقرار شود، آنگاه مراحل موجود در آن زیرمجموعه به‌عنوان مراحل مظنون در به دست آوردن انحراف تشخیص داده می‌شوند. به‌طور خلاصه، گام‌های موجود در الگوریتم تعیین مراحل مظنون به عامل انحراف به‌قرار زیرند:

گام اول: مقدار $T^۲_K$ را برای کل K متغیر موجود در فرایند محاسبه کنید. در صورتی که فرایند خارج از کنترل باشد به گام دوم بروید. در غیر این‌صورت فرایند تحت کنترل است.

گام دوم: برای زیرمجموعه‌های گروه $g_۱$ (گروهی که دارای کم‌ترین درجه‌ی آزادی است) مقدار $T^۲$ و D و D_s را محاسبه کنید. مقدار D_s را با مقدار بحرانی توزیع کای دو مقایسه کنید. اگر مقدار D_s از مقدار بحرانی بزرگ‌تر باشد، آنگاه گروه g_s (گروه مربوط به مقدار D_s) را از محاسبات بعدی کنار بگذارید و به گام سوم بروید. در غیر این‌صورت مراحل موجود در زیرمجموعه‌ی متناظر با D_s عامل انحراف به شمار می‌روند و الگوریتم به پایان می‌رسد.

گام سوم: برای گروه $g_۲$ گام دوم را تکرار کنید.

گام چهارم: برای گروه g_f گام دوم را تکرار کنید.

در صورتی‌که هیچ زیرمجموعه‌یی مظنون به عامل انحراف تشخیص داده نشد، آنگاه تمامی K مرحله، مظنون به عامل انحراف تشخیص داده می‌شوند.

تعیین متغیرهای مظنون به عامل انحراف

پس از تعیین مرحله یا مراحل مظنون به عامل انحراف، مسئله‌ی تعیین متغیر یا متغیرهایی که عامل اصلی این انحراف هستند مطرح می‌شود. در روش انتخاب زیرمجموعه‌یی از متغیرها[۴و۵]، با افزایش تعداد متغیرها تعداد تکرارهای ممکن به‌صورت توانی افزایش می‌یابد و کارایی این روش کاهش قابل توجهی می‌یابد. به طور کلی وقتی p متغیر مطرح است، تعداد حالات ممکن $۲^p - ۱$ خواهد بود. مثلاً وقتی که ۲۰ متغیر مورد بررسی است، تعداد حالات ممکن برای بررسی، ۱۰۴۸۵۷۵ حالت خواهد بود و مسلماً برای بررسی این تعداد حالات محاسبه‌گرهای بسیار قوی و صرف زمان زیاد لازم است. حال اگر بتوان شیوه‌یی ارائه داد که با استفاده از آن تعداد متغیرهای مورد بررسی را کاهش داد، صرفه‌جویی زمانی حاصل خواهد شد. به‌عبارت دیگر اگر بتوان a متغیر را از گردونه‌ی محاسبات خارج کرد، حداکثر تعداد حالات مورد بررسی برابر $۲^{p-a} - ۱$ خواهد بود.

برای دستیابی به هدف فوق می‌توان از روش تجزیه و تحلیل

مؤلفه‌های اصلی استفاده کرد. بدین ترتیب که فرض کنید بردار تصادفی $\mathbf{X}$ با متغیرهای تصادفی همبسته‌ی $X_1, X_2, ..., X_p$ و توزیع احتمال توأم $N_p(\mu,\Sigma)$ مورد بررسی قرار گیرد. اگر بردار مقادیر ویژه‌ی ماتریس واریانس ـ کوواریانس بردار تصادفی $\mathbf{X}$ به‌صورت $\lambda = [\lambda_1, \lambda_2, ..., \lambda_p]^T$ تعریف شود، آنگاه مقادیر ویژه همان ریشه‌های معادله‌یی هستند که از دترمینان $|\Sigma - \lambda I|$ به‌دست می‌آیند، و با استفاده از آنها می‌توان به ماتریس ویژه‌ی Γ دست یافت که شامل بردارهای ویژه‌ی $\gamma_1, \gamma_2, ...,$ و γ_p است.[۶] به‌عبارت دیگر می‌توان نشان داد که ماتریس متعامدی مانند Γ وجود دارد طوری که $\Gamma^T \Sigma \Gamma = D_\lambda$ و $D_\lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$. حال اگر بردار تصادفی Z به‌صورت $Z = \Gamma^T X$ تعریف شود، ماتریس واریانس ـ کوواریانس این بردار، ماتریسی است قطری که از رابطه‌ی $\Gamma^T \Sigma \Gamma = D_\lambda$ به‌دست می‌آید. به‌عبارت دیگر با استفاده از ماتریس ویژه‌ی Γ می‌توان ترکیبات خطی از متغیرهای همبسته‌ی $X_1, X_2, ...$ و X_p را طوری تعیین کرد که علاوه بر دستیابی به استقلال متغیرهای جدید که از تبدیل متغیرهای اولیه به دست می‌آیند، تغییرات به وجود آمده در متغیرهای اولیه را نیز دریافت کرد. یعنی متغیرهای مستقل جدید از طریق ترکیبات خطی متغیرهای همبسته‌ی قبل حاصل می‌شوند:

$$
\begin{aligned}
Z_1 &= \gamma_1^T \mathbf{X} = \gamma_{11} X_1 + \gamma_{12} X_2 + ... + \gamma_{1p} X_p \\
Z_2 &= \gamma_2^T \mathbf{X} = \gamma_{21} X_1 + \gamma_{22} X_2 + ... + \gamma_{2p} X_p \\
&\vdots \\
Z_p &= \gamma_p^T \mathbf{X} = \gamma_{p1} X_1 + \gamma_{p2} X_2 + ... + \gamma_{pp} X_p
\end{aligned}
\qquad (۱۳)
$$

مؤلفه‌های $Z_1, Z_2, ...$ و Z_p همان مؤلفه‌های اصلی‌اند که در آن واریانس مؤلفه‌ی Z_i برابر λ_i است. همین‌طور می‌توان نشان داد که همبستگی بین مؤلفه‌های اصلی و متغیرهای اولیه (قبل از تبدیل) از رابطه‌ی ۱۴ به دست می‌آید.

$$
\rho(X_i, Z_i) = \frac{\sqrt{\lambda_i}\,\gamma_{ii}}{\sqrt{\sigma_{ii}}} \quad ; \quad i = 1, 2, ..., p \qquad (۱۴)
$$

برای کنترل فرایند، فرض کنید Σ و در نتیجه λ و Γ در دست‌اند. نمونه‌ی جدیدی گرفته می‌شود و ماتریس واریانس ـ کوواریانس S برای این نمونه برآورد می‌شود که دارای بردار مقادیر ویژه‌ی $L = \lambda_{new}$ و ماتریس ویژه‌ی $U = \Gamma_{new}$ است. مسلماً اگر در نمونه‌ی جدید تغییراتی داده شود، این تغییرات در S و نتیجتاً در L و U منعکس می‌شود. بنابراین برای تشخیص بروز تغییرات می‌توان آزمون فرضی به صورت $H_0: U = \Gamma$ مطرح کرد که در صورت مشاهده‌ی تغییر در ماتریس مقادیر ویژه به تغییر در متغیرهای اولیه پی برد. ولی نکته‌ی اصلی در این‌است که با شناسایی تغییر در ترکیب خطی مؤلفه‌ی اصلی، چگونه می‌توان متغیر یا متغیرهای اولیه را که باعث این تغییر بوده‌اند مشخص کرد. یکی از راه‌های شناخت تعیین همبستگی بین متغیرهای اولیه و مؤلفه‌های اصلی براساس ماتریس واریانس ـ کوواریانس Σ است. بدین منظور، ماتریس همبستگی ρ را که دارای p سطر و p ستون است می‌توان از رابطه‌ی ۱۴ به‌دست آورد. واضح است که میزان همبستگی متغیر یا متغیرهای اولیه با هریک از مؤلفه‌های اصلی به ساختار ماتریس واریانس ـ کوواریانس و یا به صورت کلی به ساختار داده‌های فرایند مورد آزمون بازمی‌گردد. به‌عبارت دیگر ممکن است گاهی هریک از مؤلفه‌های اصلی بیشترین همبستگی را با تنها یک متغیر اولیه داشته باشد و گاهی نیز ممکن است تنها یک مؤلفه‌ی اصلی بیشترین همبستگی را با همه‌ی متغیرهای اولیه داشته باشد، و این همان نکته‌یی است که می‌تواند در محدودسازی فضای جواب مؤثر باشد. به بیان دیگر، در صورتی که هر یک از مؤلفه‌های اصلی تنها با یک متغیر اولیه بیشترین همبستگی را داشته باشد و تغییرات به وجود آمده ناشی از همان متغیر باشد، با بررسی بردار ویژه‌ی آن در ماتریس ویژه می‌توان متغیر مظنون را شناسایی کرد. البته این خوشبینانه‌ترین حالت است و در بدبینانه‌ترین حالت تغییر شناسایی شده در یک مؤلفه‌ی اصلی در حقیقت در همه‌ی متغیرهای اولیه وجود دارد.

برای بررسی وجود تغییر در ماتریس ویژه، فرض صفر $H_0: u_i = \gamma_i$ $(i = 1, 2, ..., p)$، را می‌توان با استفاده از آماره‌ی

$$
\Psi = N(u_i - \gamma_i)^T (\lambda_i \sum^{-1} - 2I + \lambda_i^{-1} \sum)(u_i - \gamma_i)\,; \; i = 1, 2, ..., p \qquad (۱۵)
$$

آزمود. تحت فرض صفر، آماره‌ی فوق از یک توزیع کای دو با p-۱ درجه‌ی آزادی برخوردار است. بنابراین در سطح اطمینان $(1-\alpha)$ اگر مقدار آماره‌ی فوق از مقدار بحرانی $\chi^2_{\alpha,p-1}$ بزرگتر باشد، فرض صفر رد می‌شود.

در عمل ماتریس واریانس ـ کوواریانس مشخص نیست و با استفاده از اطلاعات یک فرایند تحت کنترل برآورد می‌شود. در این حالت فرض کنید S_{ref} و a_i به‌ترتیب برآورد ماتریس واریانس ـ کوواریانس و بردارهای ویژه‌ی آن باشد. در این حالت برای بررسی تغییر در نمونه‌های جدید با N مشاهده، ماتریس واریانس ـ کوواریانس S، بردار مقادیر ویژه‌ی L، و ماتریس مقادیر ویژه‌ی U

می‌توان نشان داد که آماره‌ی مناسب برای آزمون فرض فوق به صورت زیر به دست می‌آید.

$$\Psi = N(u_i - a_i)^T (l_i S^{-1} - 2I + l_i^{-1} S)(u_i - a_i) =$$

$$N(l_i a_i^T S^{-1} a_i + l_i a_i^T S^{-1} a_i - 2); i = 1,2,...,p \qquad (16)$$

بنابراین قبل از به کارگیری روش انتخاب کلیه‌ی زیر مجموعه‌های متشکل از متغیرها (الگوریتم‌های مورفی و ختری)، با این روش قادریم که تعداد متغیرهای مورد بررسی، و در نتیجه تعداد ترکیب‌های شامل متغیرها راکاهش دهیم.

نتیجه‌گیری

در این نوشتار چگونگی آگاهی از بروز هرگونه تغییر در بردار میانگین فرایند و شناسایی عوامل اصلی این تغییر بررسی شد. برای کنترل وضعیت کل فرایند از آماره‌ی شناخته شده هاتلینگ استفاده شد و برای تعیین عوامل ایجاد شرایط عدم کنترل، مراحل تشکیل دهنده‌ی فرایند و نهایتاً متغیرهای تشکیل دهنده‌ی مراحل مورد توجه قرار گرفت.

در این پروژه، برای شناسایی مرحله یا مراحل عامل انحراف، از شیوه‌ی افراز فضای پیشامدها، که تاکنون برای شناسایی متغیر یا متغیرهای عامل انحراف به کار گرفته می‌شد، استفاده شد و نحوه‌ی استفاده از این روش به صورت یک الگوریتم ارائه شد.

پس از تعیین مرحله‌ی مظنون، شناسایی متغیر یا متغیرهای عامل انحراف مطرح است. در این راستا از ایده‌ی مؤلفه‌های اصلی و تبعیض‌گذاری کمک گرفته شد و روشی ارائه شد که با استفاده از آن می‌توان از تکرار محاسبات طولانی، که در الگوریتم‌های قبلی موجود است، کاست.

پانوشت

1. Bonferroni's Inequality
2. Bonferroni type Confidence Interval
3. chi-squar distribution

منابع

1. Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., John Wiley & Sons Inc. (1984).
2. Johnson, R.A. and Wichern, D.W., *Applied Multivariate Statistical Analysis*, 3rd Ed., Prentice Hall (1992).
3. Doganaksoy, N., Faltin, F.W. and Tucker, W.T., "Identification of out of control quality characteristics in multivariate manufacturing environment", *Communication in Statistics : Theory and Methods* **20** (9), pp 2275-2290 (1991).
4. Murphy, B.J., "Selecting out of control variables with hotelling's multivariate quality control procedure", *The Statistician,* **36** pp 571-581 (1987).
5. Srivastava, M.S. and Khatri, C.G., *An Introduction to Multivariate Statistics*, North Holland Inc. (1979).
6. Jackson, J.E., "Principal components & factor analysis : Part I - Principal components", *Journal of Quality Technology,* **12** (4), pp 201-213 (1980).
7. Ryan, T.P., *Statistical Methods for Quality Improvement*, John Wiley & Sons Inc. (1989).
8. Wade, M.R. and Woodall, W.H., "A review and analysis of cause-selecting", *Journal of Quality Technology,* **25** (3), pp 161-169 (1993).
9. Hawkins, D.M., "Multivariate quality control based on regression-adjusted variables", *Technometrics,* **33** (1), pp 61-75 (1991).
10. Hawkins, D.M., "Regression adjusted for variables in multivariate quality control", *Journal of Quality Technology,* **25** (3), pp. 170-182 (1993).
11. Healy, J.D., "A note on multivariate cusum charts", *Technometrics,* **29** (4), pp 409-412 (1987).
12. Alt, F.B., "Multivariate quality control", Encyclopedia of Statistical Sciences, **6** (1985).